The Capacity Region of the Finite-Field Multi-Way Relay Channel with Pairwise Common Messages

Lawrence Ong, Sarah J. Johnson, Christopher M. Kellett
School of Electrical Engineering and Computer Science, The University of Newcastle, Australia

***Abstract*—The capacity region of the finite-field multi-way relay channel (FFMWRC) with independent private messages was established by Ong, Johnson, and Kellett (IT-2011). In this paper, we extend the capacity results to include pairwise common messages (each known to two nodes) in addition to private messages (known to only one node). We first show that the functional-decode-forward coding scheme that achieves the capacity region of the FFMWRC with independent messages is strictly suboptimal when there are common messages. We then construct a new coding scheme that achieves the capacity region of the FFMWRC with pairwise common messages.**

I. Introduction

The multi-way relay channel models data exchange among L wireless users through a relay (e.g., base station or satellite). The *two-way* relay channel (i.e., $L = 2$) has been intensively studied [1], [2], [3]. A canonical generalization is the *multi-way* relay channel (MWRC), $L \geq 2$, where each user sends its data to all other users.

The MWRC is commonly studied in the context of independent messages [4], [5], [6], [7], [8]. Recently, we investigated the MWRC with arbitrarily correlated messages where the *uplinks* (i.e., the channels from the users to the relay) are orthogonal (i.e., non-interfering) [9]. The coding scheme proposed therein requires non-interfering uplinks. In this paper, we study the MWRC with common messages, i.e., some messages are known to two users. While common messages is a special form of correlated messages, the uplink here is interfering (the users' inputs to the channel interfere with each other), modeled by the finite-field additive channel.

We derive the capacity region of the finite-field MWRC (FFMWRC) with common messages. To this end, we construct an optimal coding scheme based on functional decode-forward [5] where the relay decodes functions of the users' messages on the uplink, and forwards the functions back to the users on the *downlink* (i.e., the channel from the relay to the users). The main difficulty is to determine an optimal function[1] that the relay should decode, taking into account that some messages are known to two users.

A. Channel Model

Fig. 1 shows the FFMWRC. Denote the users by 1, 2, ..., L, and the relay by 0. Further denote the channel input from node i by X_i, and the channel output received by node i by Y_i. The FFMWRC is defined as follows:

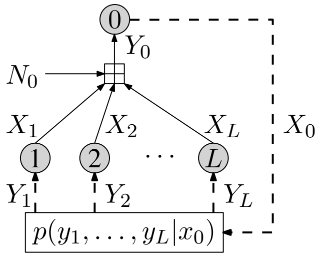

Fig. 1. The finite-field multi-way relay channel, where the solid lines represent the uplink and the dashed lines represent the downlink

$$\text{Uplink:} \quad Y_0 = X_1 \oplus X_1 \oplus \cdots \oplus X_L \oplus N_0 \triangleq \bigoplus_{i=1}^{L} X_i \oplus N_0, \tag{1}$$

$$\text{Downlink:} \quad p_{Y_1,Y_2,\ldots,Y_L|X_0}(y_1, y_2, \ldots, y_L|x_0), \tag{2}$$

where the uplink variables Y_0, N_0, and X_i for all $i \in \{1, 2, \ldots, L\}$ are elements of a finite field of order F, $\oplus$ is the addition in the field, and N_0 is arbitrarily-distributed noise. For the FFMWRC considered in this paper, we not restrict any channel model for the downlink. This general model includes the finite-field downlink channel [5] as a special case.

The finite-field model simplifies the additive white Gaussian noise (AWGN) channel—commonly used for the wireless environment—in the following ways: (i) there is no input (power) constraint for the finite-field model; and (ii) we know an optimal linear coding scheme for the multiple-access finite-field channel.[2] It nonetheless retains two important characteristics of the AWGN channel: (i) the channel is corrupted with additive noise; and (ii) the channel inputs interfere linearly with each other, i.e., the transmitted signals are added up at the receiver.[3] Coding schemes designed for the FFMWRC [4], [5] have been applied to—and shown to be optimal in—the symmetrical AWGN MWRC [11].

B. Common Messages and Rates

We use the channel n times for the users to exchange pairwise common messages in addition to private messages.

[1] Having the relay decode all the messages is strictly suboptimal in the FFMWRC [5].

[2] Though the lattice codes are optimal for the point-to-point AWGN channel, they may not be optimal for the multiple-access AWGN channel where the receiver is to decode a linear combination of the transmitters' messages [10].

[3] These are two important characteristics of multiple-access channels in general: (i) the channel is noisy, and (ii) the channel inputs interfere with each other.

The terms "private" and "common" here are defined with respect to the senders, i.e., private messages are known to only one sender, and common messages are messages known to two senders.[4] Formally, there are $L + \binom{L}{2} = \frac{L(L+1)}{2}$ independent messages denoted by $\Big\{W_{\mathcal{I}} : \mathcal{I} = \{i\}$ or $\mathcal{I} = \{i,j\}$, for all distinct $i,j \in \{1,2,\dots,L\}\Big\}$, where each $W_{\mathcal{I}} \in \{1,2,\dots,2^{nR_{\mathcal{I}}}\}$. Each user i knows its private message W_i and common messages $\Big\{W_{i,j} : j \in \{1,2,\dots,L\} \setminus \{i\}\Big\}$, where each $W_{i,j}$ is the message user i shares with user j. By definition, $W_{i,j} = W_{j,i}$. We define a rate tuple as $\boldsymbol{R} \triangleq (R_1, R_2, \dots, R_L, R_{1,2}, R_{1,3}, \dots, R_{1,L}, R_{2,3}, R_{2,4}, \dots, R_{2,L}, \dots, R_{L-1,L})$, which is a non-negative real vector of length $\frac{L(L+1)}{2}$.

We use square brackets to indicate the variable associated with a channel use. The noise term $N_0[t]$ is independent and identically distributed (iid) for each channel use $t \in \{1,2,\dots,n\}$. A length-n block code consists of the following:

1) Encoding functions at each user $i \in \{1,\dots,L\}$:
$X_i[t] = f_{it}\big(W_i, (W_{i,j})_{j\in\{1,2\dots,L\}\setminus i}, (Y_i[\tau])_{\tau\in\{1,2,\dots,t-1\}}\big)$,
for all $t \in \{1,2,\dots,n\}$.
2) Encoding functions at the relay:
$X_0[t] = f_{0t}\big((Y_0[\tau])_{\tau\in\{1,2,\dots,t-1\}}\big)$,
for all $t \in \{1,2,\dots,n\}$.
3) A decoding function at each user $i \in \{1,2,\dots,L\}$:
$\big((\hat{W}_{j(i)})_{j\in\{1,2,\dots,L\}\setminus\{i\}}, (\hat{W}_{jk(i)})_{j,k\in\{1,2,\dots,L\}\setminus\{i\}}\big)$
$= g_i\big((Y_i[\tau])_{\tau\in\{1,2,\dots,n\}}\big)$.

Here, $\hat{W}_{\mathcal{I}(i)}$ is the estimate of $\hat{W}_{\mathcal{I}}$ by user i. A decoding error occurs if some user wrongly decodes some messages, i.e., $\hat{W}_{\mathcal{I}(i)} \neq W_{\mathcal{I}}$ for some $\mathcal{I}$ and some i. This means each user needs to decode *all messages*. Assuming that each message is uniformly distributed, we denote the probability of decoding error by P_{e}. We say that the rate tuple $\boldsymbol{R}$ is achievable if the following is true: for any $\epsilon > 0$, there exists a block code such that $P_{\text{e}} \leq \epsilon$. The capacity region is the closure of all achievable rate tuples.

C. Previous Work with Only Private Messages

For the FFMWRC with only private messages (i.e., $R_{i,j} = 0$ for all (i,j)), we have the following capacity result, which is a straightforward extension of the FFMWRC with finite-field downlinks [5] (refer to Appendix A for proof):

Lemma 1: Consider the FFMWRC with only private messages, where each user i transmits an independent private message W_i. The rate $\mathbf{R}_{\text{private-only}} = (R_1, R_2, \dots, R_L)$ is achievable if there exists some $p(x_0)$ such that the following is true for all $i \in \{1,2,\dots,L\}$:

$$\sum_{j\in\{1,2,\dots,L\}\setminus\{i\}} R_j < \min\{\log_2 F - H(N_0), I(X_0;Y_i)\}. \quad (3)$$

Conversely, if $\mathbf{R}_{\text{private-only}}$ is achievable, then there exists some $p(x_0)$ such that (3) is satisfied all $i \in \{1,2,\dots,L\}$ with a non-strict inequality (i.e., $\leq$ instead of $<$).

In this paper, we refer to the coding scheme that achieves the capacity region in Lemma 1 as functional-decode-forward for private messages (FDF-P). In brief, we let two users transmit at a time and have the relay decode a function (in this case, the finite-field addition) of the messages. We cycle through different user pairs that transmit in the following order: $(1,2), (2,3), \dots, (L-1,L)$.

D. The FDF-P Scheme is Suboptimal for Common Messages

Although we can use the FDF-P scheme when there are common messages, the scheme is not always optimal. Consider an FFMWRC with three users, where

1) $X_1, X_2, X_3, Y_0, N_0 \in \{0,1,2,3\}$, i.e., $F = 4$,
2) $\Pr\{N_0 = 0\} = \Pr\{N_0 = 1\} = 1/2$,
$\Pr\{N_0 = 2\} = \Pr\{N_0 = 3\} = 0$,
3) $X_0, Y_1, Y_2, Y_3 \in \{0,1\}$,
4) $Y_1 = Y_2 = Y_3 = X_0$.

To transmit private and common messages using FDF-P, we first split each common message $W_{i,j}$ into two parts, say W_i' and W_j', and let user i transmit W_i' and user j transmits W_j' as if they were private messages. More specifically, we split the message $W_{1,2}$ into independent sub-messages, i.e., W_1' of rate R_1' and W_2' of rate R_2', where $R_1' + R_2' = R_{1,2}$. Similarly, we split the common messages (i) $W_{1,3}$ into W_1'' and W_3'' with rates R_1'' and R_3'' respectively, and (ii) $W_{2,3}$ into W_2'' and W_3' with rates R_2'' and R_3' respectively. Doing this, each user i will need to transmit (W_i, W_i', W_i'') to the other two users at the rate $r_i = R_i + R_i' + R_i''$, where $R_i', R_i'' \geq 0$.

Suppose that the following rates are achievable by FDF-P: $R_1 = R_2 = R_3 = 0.4 - \delta$, $R_{1,2} = 0.2 - \delta$, $R_{1,3} = R_{2,3} = 0.15 - \delta$ for some small $\delta > 0$. From Lemma 1, we have

$$\begin{aligned} r_1 + r_2 &= R_1 + R_2 + R_{1,2} + R_1'' + R_2'' \\ &\leq \log_2 F - H(N_0) = 1, \quad (4) \\ \Rightarrow (R_{1,3} - R_3'') &+ (R_{2,3} - R_3') = R_1'' + R_2'' \\ &\leq 1 - R_1 + R_2 + R_{1,2} = 3\delta, \quad (5) \\ \Rightarrow R_3' + R_3'' &\geq R_{1,3} + R_{2,3} - 3\delta = 0.3 - 5\delta. \quad (6) \end{aligned}$$

So, $r_1 + r_3 = R_1 + R_3 + R_3' + R_3'' + R_1' + R_1'' \geq 1.1 - 7\delta + R_1' + R_1'' \geq 1.1 - 7\delta$. Choosing $\delta = 0.01$, we have $r_1 + r_3 \geq 1.093$. But from Lemma 1, we must have $r_1 + r_3 \leq 1$ (contradiction). Hence, these rates are not achievable using the FDF-P, but they are achievable using the new scheme proposed in this paper.

The shortcoming of using FDF-P to send common messages is that the scheme ignores the fact that another user (besides the sender) knows the sub-messages, for example, user 2 knows W_1'. We will propose a nested FDF scheme that rectifies this problem.

We have previously shown that the complete-decode-forward coding scheme, where the relay decodes all the messages, is strictly suboptimal for the FFMWRC with private messages [5]. Since the FFMWRC with common messages includes that with private messages as a special case, this coding scheme is also strictly suboptimal. So, in this paper, we propose a new FDF scheme, where the relay decodes some functions of the

[4] The reader is not to be confused with common messages defined with respect to the receiver, which appear commonly in the context of broadcast channels [12], [13, p. 563].

messages. The challenge is to design optimal functions that the relay decodes.

E. Main Results

Each user a needs to decode all $\{W_\mathcal{I} : a \notin \mathcal{I}\}$. To simplify notation, we define the sum rate associated with node a as

$$R_a^\Sigma \triangleq \sum_{i \in \{1,2,\dots,L\} \setminus \{a\}} R_i + \sum_{\{i,j\} \subset \{1,2,\dots,L\} \setminus \{a\}} R_{i,j}. \quad (7)$$

We will propose a new *nested* FDF coding scheme (see Sec. III) that achieves the capacity region of the FFWMRC with common messages, given in the following theorem:

Theorem 1: Consider an FFMWRC with common messages. The rate tuple $\boldsymbol{R}$ is achievable if there exists some $p(x_0)$ such that the following is true for all $a \in \{1,2,\dots,L\}$:

$$R_a^\Sigma < \min\left\{ \log_2 F - H(N_0), I(X_0;Y_a) \right\}. \quad (8)$$

Conversely, if $\boldsymbol{R}$ is achievable, then there exists some $p(x_0)$ such that (8) holds with a non-strict inequality for all a.

Major differences between the new scheme, i.e., the nested FDF, and the existing FDF-P are as follows:

1) In FDF-P, the function that the relay decodes is predefined; in the nested FDF scheme, a function is chosen and then modified using an algorithm that we propose. As a result, the functions vary with the rates of the sub-messages. So, we might need different functions to achieve different rate tuples in the capacity region.
2) In FDF-P, the users transmit their messages, and the relay decodes functions of the messages; in the nested FDF scheme, the users transmit *functions* of their messages, and the relay decodes functions of the user's transmission (which can be functions of functions of the messages).
3) FDF-P it not always optimal when there are common messages; the nested FDF is.

II. Converse

The capacity outer bound, i.e., the converse in Theorem 1 follows directly from the following cut-set bound [13, Thm. 15.10.1]:

Lemma 2: Consider a multiterminal network with nodes $\{1,2,\dots,K\}$ where each node i sends an independent message $W^{(i\to j)}$ at the rate $R^{(i\to j)}$ to each other node j, for $j \in \{1,2,\dots,K\} \setminus \{i\}$. If the rates $\{R^{(i\to j)}\}$ are achievable, then there exists some joint probability distribution $p(x_1,x_2,\dots,x_K)$ such that

$$\sum_{i \in \mathcal{S}, j \in \mathcal{S}^\mathrm{c}} R^{(i\to j)} \leq I(X_\mathcal{S}; Y_{\mathcal{S}^\mathrm{c}} | X_{\mathcal{S}^\mathrm{c}}), \quad (9)$$

for all $\mathcal{S} \subset \{1,2,\dots,K\}$ where $X_\mathcal{S} \triangleq (X_i)_{i\in\mathcal{S}}$.

Applying Lemma 2 to the FFMWRC with $\mathcal{S} = \{1,2,\dots,\} \setminus \{a\}$ and $\mathcal{S}^\mathrm{c} = \{0,a\}$ for some $a \in \{1,2,\dots,L\}$, we have

$$\begin{aligned}
R_a^\Sigma &\leq I(X_{\{1,2,\dots,L\}\setminus\{a\}}; Y_0, Y_a | X_0, X_a) && (10a)\\
&= H(Y_0|X_0,X_a) + H(Y_a|X_0,X_a,Y_0)\\
&\quad - H(Y_0|X_{\{0,1,\dots,L\}}) - H(Y_a|X_{\{0,1,\dots,L\}},Y_0) && (10b)\\
&\leq H(Y_0) + H(Y_a|X_{\{0,1,\dots,L\}},Y_0)\\
&\quad - H(N_0) - H(Y_a|X_{\{0,1,\dots,L\}},Y_0) && (10c)\\
&= H(Y_0) - H(N_0) && (10d)\\
&\leq \log_2 F - H(N_0), && (10e)
\end{aligned}$$

where (10c) is derived as $(Y_0, X_{\{1,2,\dots,L\}}) - X_0 - Y_{\{1,2,\dots,L\}}$ forms a Markov chain, conditioning cannot increase entropy, and the channel noise N_0 is independent of the channel inputs $X_{\{0,1,\dots,L\}}$.

Similarly, setting $\mathcal{S} = \{0,1,2,\dots,\} \setminus \{a\}$ and $\mathcal{S}^\mathrm{c} = \{a\}$, we have

$$\begin{aligned}
R_a^\Sigma &\leq I(X_{\{0,1,2,\dots,L\}\setminus\{a\}}; Y_a | X_a) && (11a)\\
&= H(Y_a|X_a) - H(Y_a|X_{\{0,1,\dots,L\}}) && (11b)\\
&\leq H(Y_a) - H(Y_a|X_0) && (11c)\\
&= I(X_0;Y_a), && (11d)
\end{aligned}$$

where (11c) is derived because $(Y_0, X_{\{1,2,\dots,L\}}) - X_0 - Y_{\{1,2,\dots,L\}}$ forms a Markov chain, and as conditioning cannot increase entropy.

If the rate tuples $\boldsymbol{R}$ is achievable, then there exists some $p(x_0,x_1,\dots,x_L)$ such that (10e) and (11d) are true for all $a \in \{1,2,\dots,L\}$. Note that the RHS of (11d) depends only on the marginal pmf $p(x_0)$. This proves the converse in Theorem 1.

III. Achievability

In this section, we will prove the achievability of Theorem 1. To simplify notation, we assume (without loss of generality) that

$$R_1^\Sigma \geq R_a^\Sigma \quad (12)$$

for all $a \in \{2,3,\dots,L\}$. This means, user 1 needs to decode at the highest total rates. For any rate tuple $\boldsymbol{R}$, we can always re-index the nodes to get (12). With this simplification, we need to show that if

$$R_1^\Sigma < \log_2 F - H(N_0), \quad (13)$$

and if there exists some $p(x_0)$ such that

$$R_a^\Sigma < I(X_0;Y_a), \quad \text{for all } a \in \{1,2,\dots,L\}, \quad (14)$$

then the rate tuple $\boldsymbol{R}$ is achievable. Note that (13) is a constraint involving the uplink variables, and (14) the downlink variables.

We can think of the first constraint as the relay needing to decode at rate R_1^Σ. In general, the sum rate of all messages (i.e., $\sum_\mathcal{I} R_\mathcal{I}$) is strictly higher than R_1^Σ. This is why the complete-decode-forward coding scheme is suboptimal [5], as it requires that the relay to decode all the messages, and this imposes a constraint $\sum_\mathcal{I} R_\mathcal{I} < \log_2 F - H(N_0)$, which is stricter than (13). Our aim is to design codes such that the relay needs to

TABLE I
UPLINK MESSAGE TRANSMISSION

block	2	3	$\cdots$	L	$(2,3)$	$(2,4)$	$\cdots$	$(2,L)$	$(3,3)$	$(3,4)$	$\cdots$	$(3,L)$	$\cdots$	$(L-1,L)$
block size	k_2	k_3	$\cdots$	k_L	$k_{2,3}$	$k_{2,4}$	$\cdots$	$k_{2,L}$	$k_{3,3}$	$k_{3,4}$	$\cdots$	$k_{3,L}$	$\cdots$	$k_{L-1,L}$
row 1	$\boldsymbol{W}_2$	$\boldsymbol{W}_3$	$\cdots$	$\boldsymbol{W}_L$	$\boldsymbol{W}_{2,3}$	$\boldsymbol{W}_{2,4}$	$\cdots$	$\boldsymbol{W}_{2,L}$	$\boldsymbol{W}_{3,3}$	$\boldsymbol{W}_{3,4}$	$\cdots$	$\boldsymbol{W}_{3,L}$	$\cdots$	$\boldsymbol{W}_{L-1,L}$
row 2	$*$				$*$	$*$	$\cdots$	$*$						
row 3		$*$			$*$				$*$	$*$	$\cdots$	$*$		
$\vdots$														
row L				$*$				$*$				$*$		$*$

decode only functions of the messages, i.e., at a lower rate of R_1^Σ.[5]

A. Using Linear Codes on the Uplink

On the uplink, we will use linear block codes of the form

$$\boldsymbol{x} = (\boldsymbol{u} \odot \mathbb{G}) \oplus \boldsymbol{q}, \tag{15}$$

where the codeword $\boldsymbol{x}$, the message $\boldsymbol{u}$, and the dither $\boldsymbol{q}$ are row vectors, and $\mathbb{G}$ is the generator matrix. All elements are from the finite field of order F. In addition, each element in $\mathbb{G}$ and in $\boldsymbol{q}$ is independently and uniformly chosen over the finite field. Here, $\odot$ denotes matrix multiplication, and $\oplus$ symbol-wise addition in the finite field.

We now state a result of random linear block codes on finite-field channels [5, Thm. 3]:

Lemma 3: Consider the uplink (1). Each user encodes its message $\boldsymbol{U}_i$ (a length-k finite-field vector) to $\boldsymbol{X}_i$ (a length-n vector) using the linear code of the form (15). The users use different dithers but a common generator matrix. The receiver can decode $\bigoplus_{i=1}^L \boldsymbol{U}_i$ from its n received channel outputs $\boldsymbol{Y}_0$ with an arbitrarily small error probability if n is sufficiently large and if

$$\frac{k \log_2 F}{n} < \log_2 F - H(N_0). \tag{16}$$

B. Messages to be Transmitted on the Uplink

In order to utilize the linear block codes, we consider each message $\boldsymbol{W}_\mathcal{I}$ to be a finite-field vector of length $k_\mathcal{I}$. We have used the bold-faced symbol to emphasize that the messages are vectors. The length of the vector is thus related to the rate by $2^{nR_\mathcal{I}} = F^{k_\mathcal{I}}$. Similar to (7), we define $k_a^\Sigma \triangleq nR_a^\Sigma / \log_2 F$, i.e., the total number of finite-field symbols to be decoded by user a.

Our aim is to construct optimal functions that the relay should decode on the uplink. We will use linear codes across multiple blocks, where in each block, the relay decodes the addition of the (finite-field vector) messages transmitted in that block. Equivalently, we will determine what messages the users transmit in each block.

The transmissions are described using Table I. In each block, i.e., vertically-aligned cells from rows 1 to L, the users transmit messages assigned to the cells. Each cell contains multiple columns; each column in the cell corresponds to a message symbol (columns are not drawn in the table). We refer to the number of columns in each cell as the block size, which is set to be the message length in the cell in row 1.

We construct Table I as follows: In row 1, we place all the messages that user 1 requires, i.e., $\left\{\boldsymbol{W}_\mathcal{I} : 1 \notin \mathcal{I}\right\}$. We identify each block by the subscript, $\mathcal{I}$, of the corresponding message $\boldsymbol{W}_\mathcal{I}$ in row 1; the block size is thus $k_\mathcal{I}$.

For all the $L(L-1)/2$ messages in row 1, user $a \in \{2,3,\dots,L\}$ knows $(L-1)$ of them a priori, namely, $\Big\{\boldsymbol{W}_a, \boldsymbol{W}_{a,j} : j \in \{2,3,\dots,L\} \setminus \{a\}\Big\}$ (messages with subscript "a"). For these messages, we put an asterisk in each corresponding cell (i.e., in the same block) in row a. In these asterisked cells, we will assign messages that (i) user a requires and (ii) user 1 knows. More specifically, we will assign $\Big\{\boldsymbol{W}_1, \boldsymbol{W}_{1,j} : j \in \{2,3,\dots,L\} \setminus \{a\}\Big\}$ to these cells (replacing the each subscript "a" in row 1 by "1"). The idea is to let both users 1 and a obtain their respective required messages from the sum. These $(L-1)$ blocks in rows 1 and a are extracted out as follows:

$\boldsymbol{W}_a$	$\boldsymbol{W}_{a,2}$	$\cdots$	$\boldsymbol{W}_{a,a-1}$	$\boldsymbol{W}_{a,a+1}$	$\cdots$	$\boldsymbol{W}_{a,L}$
$\boldsymbol{W}_1, \boldsymbol{W}_{1,2}, \dots, \boldsymbol{W}_{1,a-1}, \boldsymbol{W}_{1,a+1}, \dots, \boldsymbol{W}_{1,L}$						

We spread the messages $\Big\{\boldsymbol{W}_1, \boldsymbol{W}_{1,j} : j \in \{2,3,\dots,L\} \setminus \{a\}\Big\}$ across the asterisked cells in row a (because the messages on different rows have different lengths, and they may not align at the block level). We now show that the $L(L-1)/2$ messages on row a can always fit into the corresponding $(L-1)$ asterisked cells. From (12), we have

$$\begin{aligned} &k_a + \sum_{j \in \{2,3,\dots,L\} \setminus \{a\}} (k_j + k_{a,j}) + \sum_{i,j \in \{2,3,4,\dots,L\} \setminus \{a\}} k_{i,j} \\ &\geq k_1 + \sum_{j \in \{2,3,\dots,L\} \setminus \{a\}} (k_j + k_{1,j})) + \sum_{i,j \in \{2,3,4,\dots,L\} \setminus \{a\}} k_{i,j}, \end{aligned} \tag{17}$$

which gives

$$k_a + \sum_{j \in \{2,3,\dots,L\} \setminus \{a\}} k_{a,j} \geq k_1 + \sum_{j \in \{2,3,\dots,L\} \setminus \{a\}} (k_j + k_{1,j}). \tag{18}$$

If the above equality is strict, the excess columns in the asterisked cell(s) will be left empty.

We repeat this for all $a \in \{2,3,\dots,L\}$. Doing this, in every block $\mathcal{I}$, all rows $a \in \mathcal{I}$ have asterisked cells.

[5] An additional requirement is that when the relay broadcast the function back to the users, each user can decode the messages that it requires.

C. Some Properties of Table I

We now prove a few properties of the aforementioned function construction.

Proposition 1: User 1 knows the messages in rows 2 to L a priori.

Proof: In each row a, we only assign either W_1 or $W_{1,i}$ in the asterisked cells. ■

Proposition 2: Once user a, for some $a \in \{2, 3, \dots, L\}$, knows the messages in row a, it also knows the messages in all other asterisked cells in all other rows, i.e., rows b for all $b \in \{2, 3, \dots, L\} \setminus \{a\}$.

Proof: Suppose that user a has decoded the messages in row a, i.e., $\{\boldsymbol{W}_1, \boldsymbol{W}_{1,i} : i \in \{2, 3, \dots, L\} \setminus \{a\}\}$. By definition, it knows $\boldsymbol{W}_{1,a}$ a priori. Since, any message in other rows $\{2, 3, \dots, L\} \setminus \{a\}$ must be either $\boldsymbol{W}_1$ or $\boldsymbol{W}_{1,i}$, user a knows those messages. ■

Proposition 3: For any $a \in \{1, 2, 3, \dots, L\}$, after decoding all messages in rows a and 1, user a will have decoded all the messages it requires.

Proof: Recall that user a needs to decode all messages $\Big\{\boldsymbol{W}_{\mathcal{I}} : \mathcal{I} \subset \{1, 2, \dots, L\} \setminus \{a\}, |\mathcal{I}| \leq 2\Big\}$. Clearly, row 1 contains all messages that user 1 needs to decode. Now, we prove the proposition for each $a \in \{2, 3, \dots, L\}$. For $|\mathcal{I}| = 1$, W_1 appears in row a, and $\{W_2, W_3, \dots, W_L\}$ appear in row 1. For $|\mathcal{I}| = 2$, all messages $\{\boldsymbol{W}_{i,j} : i, j \in \{2, 3, \dots, L\} \setminus \{a\}\} \triangleq \mathcal{W}_1$ appear in row 1, and $\{\boldsymbol{W}_{1,k} : k \in \{2, 3, 4, \dots, L\} \setminus \{a\}\} \triangleq \mathcal{W}_2$ in row a. So, $\mathcal{W}_1 \cup \mathcal{W}_2 = \{\boldsymbol{W}_{i,j} : i, j \in \{1, 2, \dots, L\} \setminus \{a\}\}$. ■

The following corollary is a straighforward consequence of Proposition 2:

Corollary 1: For each user $a \in \{2, 3, \dots, L\}$, there are exactly $k_1 + \sum_{i \in \{2,3,\dots,L\} \setminus \{a\}} k_{1,i}$ message *symbols* in rows 2 to L in Table I that are unknown to user a.

Proof: We know that the messages $\{\boldsymbol{W}_1, \boldsymbol{W}_{1,i} : i \in \{2, 3, \dots, L\} \setminus \{a\}\}$ are the messages in row a, and these messages are unknown to user a. The messages comprise $k_1 + \sum_{i \in \{2,3,\dots,L\} \setminus \{a\}} k_{1,i}$ message symbols. From Proposition 2, we know that there are no other messages unknown to user a in rows 2 to L. ■

D. Codewords to be Transmitted on the Uplink

With Table I constructed, we are ready to construct the uplink codewords. For each block, two users transmit—user 1 and another user $a \in \{2, 3, \dots, L\}$.

For the first $(L - 1)$ blocks, there is only one asterisk per block. For block a, for $a \in \{2, 3, \dots, L\}$, user a transmits $\boldsymbol{W}_a$ that corresponds to the cell in row 1, and simultaneously, user 1 transmits the codeword that corresponds to the asterisked cell. More specifically, user a and user 1 transmits the following respectively:

$$\boldsymbol{X}_a = (\boldsymbol{W}_a \odot \mathbb{G}) \oplus \boldsymbol{q} \tag{19}$$

$$\boldsymbol{X}_1 = (\boldsymbol{V}_a \odot \mathbb{G}) \oplus \boldsymbol{q}', \tag{20}$$

where $\boldsymbol{X}_a$, $\boldsymbol{X}_1$, $\boldsymbol{q}$, and $\boldsymbol{q}'$ each are finite-field row vectors of length n_a, $\boldsymbol{W}_a$ and $\boldsymbol{V}_a$ each are row vectors of length k_a, and $\mathbb{G}$ is an k_a-by-n_a matrix. $\boldsymbol{V}_a$ is the vector defined by the content in the asterisked cell in block a (i.e., below cell $\boldsymbol{W}_a$) in Table I. From Proposition 1, we know that user 1 knows the content in the asterisked cells, and is able to transmit the codeword (20).

For the remaining blocks $\mathcal{I} \subset \{(i, j) : i, j \in \{2, 3, \dots, L\}\}$, again two users transmit (user 1 and some user $a \in \{2, 3, \dots, L\}$) simultaneously in a similar manner. Each user encodes a length-$k_{\mathcal{I}}$ message to a length-$n_{\mathcal{I}}$ codeword as in (19)–(20) by replacing $\boldsymbol{W}_a$ with $\boldsymbol{W}_{i,j}$, and $\boldsymbol{V}_a$ with $\boldsymbol{V}_{i,j}$. As users i and j both know the message $\boldsymbol{W}_{i,j}$ in row 1, either of them transmit the codeword. User 1 transmits a function of its messages. In block (i, j), there are *two* asterisked cells—one in row i and one in row j. User 1 choose $\boldsymbol{V}_{i,j}$ as follows: if the symbols from the two asterisked cells are the same, the symbol is selected, otherwise, the finite-field sum of the symbols is selected.

From Lemma 3, if

$$\frac{k_{\mathcal{I}} \log_2 F}{n_{\mathcal{I}}} < \log_2 F - H(N_0), \tag{21}$$

and if the codelength $n_{\mathcal{I}}$ is sufficiently large, then the relay can reliably decode the function $(\boldsymbol{W}_{\mathcal{I}} \oplus \boldsymbol{V}_{\mathcal{I}})$.

Remark 1: The nested FDF scheme is *nested* in the sense that user 1 may transmit functions of messages, and hence the relay decodes *functions of functions* of the messages.

E. Decoding the Messages

Define the concatenated functions that the relay decodes as $\boldsymbol{U} \triangleq \Big(\boldsymbol{W}_{\mathcal{I}} \oplus \boldsymbol{V}_{\mathcal{I}} : \mathcal{I} \in \{2, 3, \dots, L, (i, j)_{i,j \in \{2,3,\dots,L\}}\}\Big)$ (which is a finite-field vector of length k_1^{Σ}). The relay broadcasts $\boldsymbol{U}$ on the downlink. The idea of the nested FDF scheme is that the functions $\boldsymbol{U}$ that the relay decodes and broadcasts must enable each user to obtain all its desired messages from the functions and the messages it knows a priori. To this end, we have selected the messages in the asterisked cells in Table I such that the following are true:

1) Knowing all the messages in the asterisked cells a priori (Proposition 1), user 1 can decode all messages in row 1 from $\boldsymbol{U}$, and hence obtain all its intended messages (Proposition 3).

2a) From $\boldsymbol{U}$, each user $a \in \{2, 3, \dots, L\}$ will first attempt to decode the messages in the asterisked cells in row a.

2b) Once user a decodes the messages in the asterisked cells in row a, it knows the messages in all other asterisked cells (Proposition 2). From $\boldsymbol{U}$, it can then decode the messages in row 1. With this, user a will have decoded all its intended messages (Proposition 3).

Hence, we have the following:

Proposition 4: Assuming that all users have decoded $\boldsymbol{U}$, if each user $a \in \{2, 3, \dots, L\}$ can decode the messages in the asterisked cells in row a, then all users 1 to L can decode their intended messages.

The decoding procedure now hinges on Step 2a. Let us focus on one user a. User a attempts to decode the messages in the asterisked cells in row a, denoted by

$\boldsymbol{W}_\Lambda \triangleq \big\{\boldsymbol{W}_1, \boldsymbol{W}_{1,i} : i \in \{2,3,\dots,L\} \setminus \{a\}\big\}$. It does so from the relevant parts of $\boldsymbol{U}$, namely, $\{\boldsymbol{W}_\mathcal{I} \oplus \boldsymbol{V}_\mathcal{I} : \mathcal{I} \in \Theta\}$, where $\Theta \triangleq \{a, (a,i) : i \in \{2,3,\dots,L\} \setminus \{a\}\}$ are all the blocks with asterisked cells in row a. Recall that each symbol in $\boldsymbol{V}_\Theta \triangleq \{\boldsymbol{V}_\mathcal{I} : \mathcal{I} \in \Theta\}$ is the sum of distinct symbols in a column from rows 2 to L. The blocks in Θ are shown in the table below:

row 1	$\boldsymbol{W}_a$	$\boldsymbol{W}_{a,2}$	$\boldsymbol{W}_{a,3}$	$\cdots$	$\boldsymbol{W}_{a,L}$	$\boldsymbol{W}_\Theta$	
row 2		$*$		$\cdots$			$\boldsymbol{V}_\Theta$
$\vdots$							
row a	$\boldsymbol{W}_1, \boldsymbol{W}_{1,2}, \dots, \boldsymbol{W}_{1,L}$ (all $*$ cells in row a)					$\boldsymbol{W}_\Lambda$	
$\vdots$							
row L				$\cdots$	$*$		

As user a knows $\boldsymbol{W}_\Theta$, it subtracts these from $\boldsymbol{W}_\Theta \oplus \boldsymbol{V}_\Theta$ to obtain $\boldsymbol{V}_\Theta$. Note that $\boldsymbol{W}_\Lambda$ comprises $k_1 + \sum_{i\in\{2,3,\dots,L\}\setminus\{a\}} k_{1,i} \triangleq A'$ symbols. Denote the first A' symbols of $\boldsymbol{V}_\Theta$ by $\boldsymbol{V}'_\Theta$, which are functions of only $\boldsymbol{W}_\Lambda$ (Proposition 2). So, Step 2a is successful if and only if $\boldsymbol{V}'_\Theta$ form A' linearly independent equations. We now propose an algorithm to rearrange the messages in the asterisked cells to achieve this.

F. An Algorithm to Shuffle the Messages

Our aim is to get $A' \triangleq k_1 + \sum_{i\in\{2,3,\dots,L\}\setminus\{a\}} k_{1,i}$ linearly independent equations in $\boldsymbol{V}'_\Theta$ (*simultaneously* for all users 2 to L). Hence, we can ignore row 1; we only need to consider asterisked cells in Table I. Each block has at most two asterisked cells. For each column from row 2 to row L, we denote the *simplified column* by $\begin{bmatrix}\alpha\\\beta\end{bmatrix}$, where α and β are symbols from the asterisked cells. If there is only one asterisk cell for that column, we have $\begin{bmatrix}\alpha\end{bmatrix}$. Using this notation, we now propose the shuffling algorithm.

```
repeat
    foreach user $a = 2, 3, \dots, L$ do
        Consider all the non-empty columns in row $a$;
        Rearrange each simplified column such that the
        top element takes the symbol in row $a$;
        while There exists two columns $\begin{bmatrix}\alpha\\\beta\end{bmatrix}$ and $\begin{bmatrix}\gamma\\\alpha\end{bmatrix}$ for
        some $\beta \neq \alpha$ and $\gamma \neq \alpha$ do
            Swap $\alpha$ and $\gamma$ to get $\begin{bmatrix}\gamma\\\beta\end{bmatrix}$ and $\begin{bmatrix}\alpha\\\alpha\end{bmatrix}$;
        end
    end
until The while condition is not satisfied for one cycle of
the foreach loop;
```

Algorithm 1: The Shuffling Algorithm

Although swapping the elements for one user may affect the simplified columns for other users, in each swap, we always increase the number of simplified columns with two identical elements, i.e., $\begin{bmatrix}\alpha\\\alpha\end{bmatrix}$. As there are only a finite number of columns, the algorithm will always terminate after at most k_1^Σ cycles of the **foreach** loop.

We will now show that when the algorithm ends, user $a \in \{2,3,\dots,L\}$ can decode the messages on row a, denoted by $\boldsymbol{W}_\Lambda$ (consisting of A' symbols). We treat $\begin{bmatrix}\alpha\\\alpha\end{bmatrix}$ as $\begin{bmatrix}\alpha\end{bmatrix}$ because only one copy of the same symbol is transmitted to the relay. When we swap the elements for each user in the algorithm, we always swap the top elements, i.e., symbols in the same row—the bottom elements may be from different rows.

Consider the decoding of user a using the A' non-empty simplified columns. The user again rearranges each simplified column such that the top element takes the symbol in row a. Doing that, the top elements of these simplified columns are distinct—they corresponds to the A' symbols in $\boldsymbol{W}_\Lambda$. Now, if a symbol α appear at the top of a simplified column and the bottom of another, it has to take the form $\begin{bmatrix}\alpha\end{bmatrix}, \begin{bmatrix}\beta\\\alpha\end{bmatrix}, \begin{bmatrix}\gamma\\\alpha\end{bmatrix}, \dots, \begin{bmatrix}\zeta\\\alpha\end{bmatrix}$, because all α can only appear once at the top, and the case $\begin{bmatrix}\alpha\\\beta\end{bmatrix}$ and $\begin{bmatrix}\gamma\\\alpha\end{bmatrix}$ for any $\beta \neq \alpha$ and $\gamma \neq \alpha$ has been eliminated. Consequently, user a can decode $\boldsymbol{W}_\Lambda$. This is true for all $a \in \{2,3,\dots,L\}$. Combining this with Proposition 4, we have the following:

Proposition 5: If a user can decode $\boldsymbol{U}$, then it can decode all its intended messages.

G. Rate Bounds for Decoding $\boldsymbol{U}$

Now, we derive the rates at which the users can transmit such that all users can reliably decode $\boldsymbol{U}$.

We first determine the condition for which the relay can reliably decode $\boldsymbol{U}$. On the uplink, the users transmit k_1^Σ aligned symbols in n channel uses across all blocks in Table I. We allocate the number of channel uses for each block proportional to the message length, i.e.,

$$n_\mathcal{I} = n\frac{k_\mathcal{I}}{k_1^\Sigma}, \tag{22}$$

for all $\mathcal{I} \in \{2,3,\dots,L,(i,j)_{i,j\in\{2,3,\dots,L\}}\}$. If

$$R_1^\Sigma = \frac{k_1^\Sigma \log_2 F}{n} = \frac{k_\mathcal{I} \log_2 F}{n_\mathcal{I}} < \log_2 F - H(N_0), \tag{23}$$

and if n is sufficiently large, then (21) is satisfied for all $\mathcal{I}$, and thus the relay can reliably decode $\boldsymbol{U}$.

Now, suppose that the relay has successfully decoded $\boldsymbol{U}$. Note that $\boldsymbol{U} \triangleq (U^{(1)}, U^{(2)}, \dots, U^{(k_1^\Sigma)})$ is a finite-field vector of length k_1^Σ. Hence, there are at most $F^{k_1^\Sigma}$ distinct vectors $\boldsymbol{U}$. On the downlink, the relay first chooses some $p(x_0)$, generates $F^{k_1^\Sigma}$ sequences $\boldsymbol{x_0}$, each of length n, and indexes them as $\boldsymbol{X}_0(u^{(1)}, u^{(2)}, \dots, u^{(k_1^\Sigma)}) = \boldsymbol{X}_0(\boldsymbol{u})$. Upon decoding $\boldsymbol{U}$ on the uplink, the relay transmits $\boldsymbol{X}_0(\boldsymbol{U})$ on the downlink.

Each user $a \in \{1,2,\dots,L\}$ attempts to decode $\boldsymbol{U}$ on the downlink with the help of its prior messages, $\{\boldsymbol{W}_\mathcal{I} : a \in \mathcal{I}\}$. With the correct prior messages, all ambiguities in $\boldsymbol{U}$ can only be caused by different $\{\boldsymbol{W}_\mathcal{I} : a \notin \mathcal{I}\}$. For user a, let $\mathcal{D}_a$ denotes the set of distinct $\boldsymbol{u}$ that can be formed by all possible $\{\boldsymbol{W}_\mathcal{I} : a \notin \mathcal{I}\}$. We have

$$|\mathcal{D}_a| \leq |\{\boldsymbol{W}_\mathcal{I} : a \notin \mathcal{I}\}| = F^{k_a^\Sigma}. \tag{24}$$

The set $\mathcal{D}_a$ contains all possible $\boldsymbol{u}$ user a may decode to.

Knowing the messages $\{\boldsymbol{W}_{\mathcal{I}} : a \in \mathcal{I}\}$, each user $a \in \{1, 2, \ldots, L\}$ decodes $\boldsymbol{U}$ from its received downlink channel outputs $\boldsymbol{Y}_a$ if it can find a unique vector $\boldsymbol{u} \in \mathcal{D}_a$ such that

$$(\boldsymbol{X}_0(\boldsymbol{u}), \boldsymbol{Y}_a) \in A_\epsilon^{(n)}(X_0, Y_a), \tag{25}$$

where $A_\epsilon^{(n)}(X_0, Y_a)$ is the set of jointly typical sequences $\{(\boldsymbol{x}_0, \boldsymbol{y}_a)\}$ [13, p. 195]. Otherwise, user a declares a decoding error. So, user a makes an error in decoding if the event $E_1 \cup E_2$ occurs, where

- E_1: the correct $\boldsymbol{U} \in \mathcal{D}_a$ does not satisfy (25),
- E_2: some wrong $\boldsymbol{u}' \in \mathcal{D}_a$ satisfies (25).

By definition, $\boldsymbol{U} \in \mathcal{D}_a$. From the joint asymptotic equipartition property (JAEP), we know that [13, Thm. 7.6.1]

$$\Pr\{E_1\} < \epsilon. \tag{26}$$

We now evaluate the probability of E_2:

$$\begin{aligned} \Pr\{E_2\} &= \Pr\{\text{some } \boldsymbol{u}' \in \mathcal{D}_a \text{ satisfies (25)}\} && \text{(27a)} \\ &\leq \sum_{\boldsymbol{u}' \in \mathcal{D}_a \setminus \{\boldsymbol{U}\}} \Pr\{\boldsymbol{u}' \text{ satisfies (25)}\} && \text{(27b)} \\ &\leq (F^{k_a^\Sigma} - 1) 2^{-n(I(X_0;Y_a) - 3\epsilon)} && \text{(27c)} \\ &< 2^{n([k_a^\Sigma \log_2 F]/n - I(X_0;Y_a) + 3\epsilon)}, && \text{(27d)} \end{aligned}$$

where (27c) follows from (24) and the JAEP [13, Thm. 7.6.1].

This means by choosing a sufficiently small ϵ and a sufficiently large n, if

$$R_a^\Sigma = \frac{k_a^\Sigma \log_2 F}{n} < I(X_0; Y_a), \tag{28}$$

then $\Pr\{E_1 \cup E_2\} \leq \Pr\{E_1\} + \Pr\{E_2\}$ can be made arbitrarily small, meaning that user a can realibly decode $\boldsymbol{U}$.

Consequently, if (13) and (14) are satisfied for all $a \in \{1, 2, \ldots, L\}$, then all users can reliably decode $\boldsymbol{U}$. It follows from Proposition 4 that each user can reliably decode its intended messages. With this, we have proven the achievability of Theorem 1.

Remark 2: In our proposed coding scheme, the relay transmits after decoding $\boldsymbol{U}$. This means a total of $2n$ channel uses (n for the uplink, followed by n for the downlink). This issue can be easily fixed by repeating this scheme for multiple messages over multiple *blocks*. Using the uplink and the downlink simultaneously, the relay transmits $\boldsymbol{U}$ that it has previously decoded in the previous block, while, at the same time, the users transmits new messages. This is a commonly-used technique for relay channels (see, e.g., [14][5]).

Appendix A
Generalization of the Capacity Results for Arbitrary Downlink

We have previously established [5, Sec. V.A] that if

$$\sum_{j \in \{1,2,\ldots,L\} \setminus \{i\}} R_j < \log_2 F - H(N_0), \tag{29}$$

for all $i \in \{1, 2, \ldots, L\}$, then the relay can reliably decode a function (denoted by $\boldsymbol{U}$) of the users' messages. From [5, eqns. (59),(60),(63f),(64e)], we know that if

$$\sum_{j \in \{1,2,\ldots,L\} \setminus \{i\}} R_j < I(X_0; Y_i), \tag{30}$$

for all $i \in \{1, 2, \ldots, L\}$ and for some $p(x_0)$, then each user i can reliably decode $\boldsymbol{U}$, and subsequently obtain the other users' messages. This proves the achievability of rates satisfying (3). The converse follows from Section II by setting all $R_{i,j} = 0$.

References

[1] B. Rankov and A. Wittneben, "Achievable rate regions for the two-way relay channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Seattle, USA, July 9–14 2006, pp. 1668–1672.

[2] W. Nam, S. Chung, and Y. H. Lee, "Capacity of the Gaussian two-way relay channel to within $\frac{1}{2}$ bit," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5488–5494, Nov. 2010.

[3] M. P. Wilson, K. Narayanan, H. D. Pfister, and A. Sprintson, "Joint physical layer coding and network coding for bidirectional relaying," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5641–5654, Nov. 2010.

[4] L. Ong, S. J. Johnson, and C. M. Kellett, "An optimal coding strategy for the binary multi-way relay channel," *IEEE Commun. Lett.*, vol. 14, no. 4, pp. 330–332, Apr. 2010.

[5] ——, "The capacity region of multiway relay channels over finite fields with full data exchange," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3016–3031, May 2011.

[6] C. Hausl, O. Işcan, and F. Rossetto, "Resource allocation for asymmetric multi-way relay communication over orthogonal channels," *EURASIP J. Wirel. Commun.*, vol. 2012, no. 20, Jan. 2012.

[7] T. Cui, J. Kliewer, and T. Ho, "Communication protocols for N-way all-cast relay networks," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3239–3251, Nov. 2012.

[8] D. Gündüz, A. Yener, A. Goldsmith, and H. V. Poor, "The multiway relay channel," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 51–63, Jan. 2013.

[9] R. Timo, G. Lechner, L. Ong, and S. J. Johnson, "Multi-way relay networks: Orthogonal uplink, source-channel separation and code design," *to appear in IEEE Trans. Commun.*, Feb., 2013. [Online]. Available: http://arxiv.org/pdf/1210.0271v1.pdf

[10] U. Erez and R. Zamir, "A modulo-lattice transformation for multiple-access channels," in *Proc. 25th IEEE Conv. Electr. Electron. Eng. Israel*, Tel Aviv, Israel, Dec. 3–5 2008, pp. 836–840.

[11] L. Ong, C. M. Kellett, and S. J. Johnson, "On the equal-rate capacity of the AWGN multiway relay channel," *IEEE Trans. Inf. Theory*, vol. 58, no. 9, pp. 5761–5769, Sept. 2012.

[12] R. F. Wyrembelski, T. J. Oechtering, and H. Boche, "MIMO Gaussian bidirectional broadcast channels with common messages," *IEEE Trans. Wirel. Commun.*, vol. 10, no. 9, pp. 2950–2959, Sept. 2011.

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, 2006.

[14] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 5, pp. 572–584, Sept. 1979.